

**Table W1.** Frequency Analysis of Gene Occurrence on Training Data Sets.

A. Frequency of Gene Occurrences in the 2000 Classifiers for the Khan et al. SRBCT Study

Index	Clone_ID	Symbol	Z Score
1	784224	<i>FGFR4</i>	52.17
122	812105	<i>AF1Q</i>	48.16
128	814260	<i>FVT1</i>	48.16
173	796258	<i>SGCA</i>	46.43
186	1435862	<i>CD99</i>	45.29
245	770394	<i>FCGRT</i>	45.29
254	377461	<i>CAV1</i>	43.57
367	207274	<i>NULL</i>	42.99
421	866702	<i>PTPN13</i>	39.55
482	244618	<i>FRCP2</i>	39.55
508	298062	<i>TNNT2</i>	34.96
544	52076	<i>OLFM1</i>	33.82
553	769716	<i>NF2</i>	32.67
565	491565	<i>CITED2</i>	31.52
602	325182	<i>CDH2</i>	30.95
606	296448	<i>IGF2</i>	29.80
741	143306	<i>NULL</i>	28.65
822	25725	<i>FDFT1</i>	28.08
835	308231	<i>MYO1B</i>	27.51
845	43733	<i>GYG2</i>	25.79
909	236282	<i>WAS</i>	25.21
975	859359	<i>TP53I3</i>	25.21
1002	1471841	<i>ATP1A1</i>	24.64
1054	839552	<i>NCOA1</i>	24.07
1065	461425	<i>MYL4</i>	22.92
1073	629896	<i>MAP1B</i>	22.92
1083	486110	<i>PFN2</i>	22.34
1104	788107	<i>BIN1</i>	22.34
1193	42558	<i>GATM</i>	21.20
1206	241412	<i>ELF1</i>	21.20
1318	713922	<i>GSTM2</i>	21.20
1326	878652	<i>PCOLCE</i>	20.62
1388	786084	<i>CBX1</i>	20.05
1426	1409509	<i>TNNT1</i>	19.48
1600	244637	<i>CRI1</i>	19.48
1644	789253	<i>PSEN2</i>	18.33
1661	357031	<i>TNFAIP6</i>	17.76
1707	134748	<i>GCSH</i>	17.76
1722	183337	<i>HLA-DMA</i>	17.76
1737	813841	<i>PLAT</i>	17.76
1771	814444	<i>CRSP9</i>	17.18
1798	377048	<i>MYO1B</i>	17.18
1885	771323	<i>PLOD</i>	17.18
1895	897788	<i>PTPRF</i>	17.18
1910	214572	<i>CDK6</i>	16.61
1953	142134	<i>ARHGAP8</i>	16.61
1954	756401	<i>RHEB</i>	16.03
1979	293500	<i>FLJ90440</i>	16.03
2045	811108	<i>TRIP6</i>	15.46
2049	504791	<i>GSTA4</i>	15.46
2143	1473131	<i>TLE2</i>	14.89
2145	898219	<i>MEST</i>	14.89
2156	841641	<i>CCND1</i>	14.89
2158	295985	<i>NULL</i>	14.89

B. Frequency of Gene Occurrences in the 1000 Classifiers for the Latulippe et al. Prostate Study

Probe Set	Symbol	Z Score
37407_s_at	<i>MYH11</i>	147.59
32582_at	<i>MYH11</i>	133.02
767_at	<i>MYH11</i>	126.4
1197_at	<i>ACTG2</i>	122.42
36931_at	<i>TAGLN</i>	101.23
32755_at	<i>ACTA2</i>	85.34
37576_at	<i>PCP4</i>	81.36
774_g_at	<i>MYH11</i>	80.04
34203_at	<i>CNN1</i>	40.3

**Table W1.** (continued)

B. Frequency of Gene Occurrences in the 1000 Classifiers for the Latulippe et al. Prostate Study

Probe Set	Symbol	Z Score
36834_at	<i>MOXD1</i>	36.33
39333_at	<i>COL4A1</i>	35.01
773_at	<i>MYH11</i>	31.03
38834_at	<i>TOPBP1</i>	29.71
685_f_at	<i>LOC112714</i>	29.71
34878_at	<i>SMC4L1</i>	28.38
35970_g_at	<i>MPHOSPH9</i>	28.38
41137_at	<i>PPP1R12B</i>	27.06
1884_s_at	<i>PCNA</i>	25.73
40407_at	<i>KPNA2</i>	25.73
32662_at	<i>MDC1</i>	24.41
34376_at	<i>PKIG</i>	24.41
35742_at	<i>C16orf45</i>	24.41
36987_at	<i>LMNB2</i>	24.41
39145_at	<i>MYL9</i>	23.09
38430_at	<i>FABP4</i>	21.76
1599_at	<i>CDKN3</i>	20.44
2012_s_at	<i>PRKDC</i>	20.44
32305_at	<i>COL1A2</i>	20.44
418_at	<i>MKI67</i>	20.44
651_at	<i>RPA3</i>	20.44
35474_s_at	<i>COL1A1</i>	19.11
37749_at	<i>MEST</i>	19.11
38031_at	<i>DDX48</i>	19.11
39990_at	<i>ISL1</i>	19.11
1505_at	<i>TYMS</i>	17.79
33924_at	<i>RAB6IP1</i>	17.79
35694_at	<i>MAP4K4</i>	17.79
32306_g_at	<i>COL1A2</i>	16.46
32847_at	<i>MYLK</i>	16.46
37305_at	<i>EZH2</i>	16.46
41081_at	<i>BUB1</i>	16.46
32272_at	<i>K-ALPHA-1</i>	15.14
36627_at	<i>SPARCL1</i>	15.14
37347_at	<i>CKS1B</i>	15.14
39519_at	<i>KIAA0692</i>	15.14
40845_at	<i>ILF3</i>	15.14

C. Frequency of Gene Occurrences in the 2000 Classifiers for the Lapointe et al. Prostate Cancer Study

Index	CLID	Symbol	Z Score
1590	IMAGE:288663	<i>GJB1</i>	282.87
250	IMAGE:136605	<i>POMP</i>	264.49
3388	IMAGE:788180	<i>AMACR</i>	225.69
812	IMAGE:179211	<i>GPR160</i>	222.63
44	IMAGE:1034473	<i>AMACR</i>	205.27
1094	IMAGE:2057931	<i>TACSTD1</i>	182.81
2817	IMAGE:685516	<i>GPR160</i>	174.64
2330	IMAGE:470216	<i>MYO6</i>	154.22
2072	IMAGE:396839	<i>NULL</i>	118.48
782	IMAGE:1709503	<i>NULL</i>	96.02
2109	IMAGE:415962	<i>PCSK6</i>	86.83
3017	IMAGE:745283	<i>DKFZp779Q175</i>	86.83
1592	IMAGE:288770	<i>TMEM144</i>	76.62
1098	IMAGE:2062404	<i>MON1B</i>	68.45
1295	IMAGE:244391	<i>PRKCBP1</i>	68.45
2277	IMAGE:460164	<i>C20orf74</i>	64.37
4104	IMAGE:898286	<i>CDC2</i>	54.16
1076	IMAGE:2043415	<i>DAPK1</i>	52.11
3951	IMAGE:855406	<i>NULL</i>	52.11
2116	IMAGE:416374	<i>TSPAN13</i>	52.11
2583	IMAGE:510534	<i>TACSTD1</i>	51.09
1294	IMAGE:244350	<i>NULL</i>	48.03
2589	IMAGE:510856	<i>ENC1</i>	47.01
3580	IMAGE:811101	<i>SH3RF1</i>	43.95
1249	IMAGE:23819	<i>ABCG1</i>	42.92
1173	IMAGE:2244718	<i>GCNT1</i>	40.16

**Table W2.** Classifiers Generated by GP for Each Training Data Set.

Classification Rules	Gene Description
<b>A. Classifiers that distinguish different cancer classes of SRBCT</b>	
IF IMAGE:183337 > exp(IMAGE:796613) THEN BL	IMAGE:183337 is HLA-DMB; IMAGE:796613 is COL5A2;
IF (IMAGE:767183 – IMAGE:504791) > IMAGE:813742 THEN BL	IMAGE:767183 is HCLS1; IMAGE:504791 is GSTA4; IMAGE:813742 is XPO6;
IF (IMAGE:767183/IMAGE:810408 – IMAGE:504791) > IMAGE:809383 THEN BL	IMAGE:767183 is HCLS1; IMAGE:810408 is ERGIC3; IMAGE:504791 is GSTA4; IMAGE:809383 is PPNAN;
IF log(sqrt(IMAGE:183337)) > tanh(IMAGE:308746) THEN BL	IMAGE:183337 is HLA-DMB; IMAGE:308746 is EXT2;
IF (IMAGE:740604*IMAGE:183462 – IMAGE:250654) > (IMAGE:325155 * IMAGE:813444)*IMAGE:815555 THEN BL	IMAGE:740604 is ISG20; IMAGE:183462 is MAN2C1; IMAGE:250654 is SPARC; IMAGE:325155 is KRT34; IMAGE:813444 is SLC12A8; IMAGE:815555 is DGKA;
IF (IMAGE:767183 – IMAGE:504791) >= IMAGE:813742 THEN BL	IMAGE:767183 is HCLS1; IMAGE:504791 is GSTA4; IMAGE:813742 is XPO6;
IF (sin(IMAGE:140171) > IMAGE:854899^2) && ((IMAGE:767183 – IMAGE:504791) > IMAGE:486110^2) THEN BL	IMAGE:140171 is THTPA; IMAGE:854899 is DUSP6; IMAGE:767183 is HCLS1; IMAGE:504791 is GSTA4; IMAGE:486110 is PFN2;
IF (IMAGE:740604*cos(IMAGE:897632) – IMAGE:250654) > IMAGE:183462 THEN BL	IMAGE:740604 is ISG20; IMAGE:897632 is ATP2A2; IMAGE:250654 is SPARC; IMAGE:183462 is MAN2C1;
IF tanh(IMAGE:203003) <= IMAGE:740604 THEN BL	IMAGE:203003 is DECR2; IMAGE:740604 is ISG20;
IF (IMAGE:740604*tanh(IMAGE:563673)) > (2.0 * IMAGE:250654) THEN BL	IMAGE:740604 is ISG20; IMAGE:563673 is ALDH7A1; IMAGE:250654 is SPARC
IF (IMAGE:866702 * IMAGE:814260) > IMAGE:295985 THEN EWS	IMAGE:866702 is PTPN13; IMAGE:814260 is FVT1; IMAGE:295985 is CDK6;
IF (IMAGE:866702/IMAGE:1469230) > IMAGE:295985 THEN EWS	IMAGE:866702 is PTPN13; IMAGE:1469230 is COX8A; IMAGE:295985 is CDK6;
IF sqrt(IMAGE:770394/IMAGE:52076) <= IMAGE:770394 THEN EWS	IMAGE:770394 is FCGRT; IMAGE:52076 is OLFM1; IMAGE:770394 is FCGRT;
IF (IMAGE:178825 * IMAGE:52076) > IMAGE:295985 THEN EWS	IMAGE:178825 is NRGN; IMAGE:52076 is OLFM1; IMAGE:295985 is CDK6;
IF (IMAGE:866702 * IMAGE:80338) > IMAGE:295985 THEN EWS	IMAGE:866702 is PTPN13; IMAGE:80338 is SELENBP1; IMAGE:295985 is CDK6;
IF IMAGE:770394 > (sqrt(IMAGE:244951))/(IMAGE:357031/IMAGE:855391)) THEN EWS	IMAGE:770394 is FCGRT; IMAGE:244951 is HIF3A; IMAGE:357031 is TNFAIP6; IMAGE:855391 is STK25;
IF IMAGE:43733 > IMAGE:295985 THEN EWS	IMAGE:43733 is GYG2; IMAGE:295985 is CDK6;
IF (IMAGE:770394 + IMAGE:214906) > IMAGE:842820 THEN EWS	IMAGE:770394 is FCGRT; IMAGE:214906 is SPEN; IMAGE:842820 is PABPC4;
IF IMAGE:866702 * sqrt(IMAGE:138550) > IMAGE:295985 THEN EWS	IMAGE:866702 is PTPN13; IMAGE:138550 is ZNF444; IMAGE:295985 is CDK6;
IF IMAGE:377461 > sqrt(IMAGE:132911) THEN EWS	IMAGE:377461 is CAV1; IMAGE:132911 is PPP1CB;
IF IMAGE:364510 > IMAGE:810057^2 THEN NB	IMAGE:364510 is SATB1; IMAGE:810057 is CSDA;
IF (3.951/sqrt(IMAGE:1409509)) > (exp(IMAGE:810057^2)/(IMAGE:135688^2 + sqrt(IMAGE:1472775)^2) THEN NB	IMAGE:1409509 is TNNT1; IMAGE:810057 is CSDA; IMAGE:135688 is GATA2; IMAGE:1472775 is COL8A1;
IF ((8.714/exp(IMAGE:121521))/exp(IMAGE:546600)) > ((IMAGE:810057^2)/(IMAGE:325182^2)) THEN NB	IMAGE:121521 is ZNF764; IMAGE:546600 is DNAJB4; IMAGE:810057 is CSDA; IMAGE:325182 is CDH2;
IF IMAGE:364510 > (IMAGE:810057 + ((IMAGE:759200^2)/(4.276/sin(IMAGE:24415)))) THEN NB	IMAGE:364510 is SATB1; IMAGE:810057 is CSDA; IMAGE:759200 is DHPS; IMAGE:24415 is TP53;
IF IMAGE:811956 > IMAGE:810057 AND (IMAGE:23132 > cos(IMAGE:491001)) THEN NB	IMAGE:811956 is RAN; IMAGE:810057 is CSDA; IMAGE:23132 is SF3A1; IMAGE:491001 is GLO1;
IF IMAGE:325182*((exp(IMAGE:840942))^2) > ((IMAGE:789376*(-1.222))/1.791) THEN NB	IMAGE:325182 is CDH2; IMAGE:840942 is HLA-DPB1; IMAGE:789376 is TXNRD1;
IF IMAGE:823886 > exp(IMAGE:810057) THEN NB	IMAGE:823886 is NDEL1; IMAGE:810057 is CSDA;
IF ((exp(IMAGE:812105)/5.506)^2) > (IMAGE:323577 + IMAGE:824426 * IMAGE:814266) THEN NB	IMAGE:812105 is MLLT11; IMAGE:323577 is SLC4A2; IMAGE:824426 is PDAP1; IMAGE:814266 is PRKCZ;
IF exp(IMAGE:810057) <= (IMAGE:395708 + IMAGE:788205 + IMAGE:383188) THEN NB	IMAGE:810057 is CSDA; IMAGE:395708 is DPYSL4; IMAGE:788205 is SOX4; IMAGE:383188 is RCVRN;
IF IMAGE:810057 <= IMAGE:29054 THEN NB	IMAGE:810057 is CSDA; IMAGE:29054 is SCGN;
IF (IMAGE:298062/IMAGE:40017) > (IMAGE:549146*IMAGE:784224) THEN RMS	IMAGE:298062 is TNNT2; IMAGE:40017 is CYCS; IMAGE:549146 is TRIM22; IMAGE:784224 is FGFR4;
IF (IMAGE:511909/IMAGE:784224) <= IMAGE:461425 THEN RMS	IMAGE:511909 is CDH17; IMAGE:784224 is FGFR4; IMAGE:461425 is MYL4;
IF (IMAGE:898219^2 + IMAGE:461425)^2 > IMAGE:82131 THEN RMS	IMAGE:898219 is MEST; IMAGE:461425 is MYL4; IMAGE:82131 is PSMB1;
IF ((IMAGE:265102 + IMAGE:207274) > IMAGE:244618) AND ((IMAGE:784224 + IMAGE:244618) > 1.8594) THEN RMS	IMAGE:265102 is ABI2; IMAGE:207274 is IGF2; IMAGE:244618 is FNDC5; IMAGE:784224 is FGFR4; IMAGE:244618 is FNDC5;
IF IMAGE:796258 > (IMAGE:340630/IMAGE:784224) THEN RMS	IMAGE:796258 is SGCA; IMAGE:340630 is MAP3K8; IMAGE:784224 is FGFR4;
IF (IMAGE:244618/IMAGE:52076) > (IMAGE:809939/IMAGE:813841) THEN RMS	IMAGE:244618 is FNDC5; IMAGE:52076 is OLFM1; IMAGE:809939 is C6orf48; IMAGE:813841 is PLAT;
IF (IMAGE:950367/(IMAGE:244618/IMAGE:52076)) <= (IMAGE:142134 + IMAGE:143306^2) THEN RMS	IMAGE:950367 is ADAR; IMAGE:244618 is FNDC5; IMAGE:52076 is OLFM1; IMAGE:142134 is PRR5; IMAGE:143306 is LOC645166;

**Table W2.** (continued)

Classification Rules	Gene Description
<b>A. Classifiers that distinguish different cancer classes of SRBCT</b>	
IF IMAGE:784224 > IMAGE:781341 THEN RMS	IMAGE:784224 is FGFR4; IMAGE:781341 is TULP4;
IF (IMAGE:461425 + IMAGE:898219 <sup>2</sup> ) > IMAGE:825312 THEN RMS	IMAGE:461425 is MYL4; IMAGE:898219 is MEST; IMAGE:825312 is ATP5J;
IF IMAGE:298062 > (0.045022 + IMAGE:110467 + sin(IMAGE:784224)) THEN RMS	IMAGE:298062 is TNNT2; IMAGE:110467 is CAV2; IMAGE:784224 is FGFR4;
<b>B. Classifiers that distinguish metastatic prostate cancer from primary prostate cancer</b>	
IF (36987_at >= 767_at) THEN MET	36987_at is LMNB2; 767_at is MYH11
IF (40145_at >= 32582_at) THEN MET	40145_at is TOP2A; 32582_at is MYH11
IF (40398_s_at > 32755_at) THEN MET	40398_s_at is MEOX2; 32755_at is ACTA2
IF (39145_at <= 39333_at) THEN MET	39145_at is MYL9; 39333_at is COL4A1
IF (767_at < 37263_at) THEN MET	767_at is MYH11; 37263_at is GGH
IF (38505_at >= 37576_at) THEN MET	38505_at is EIF2C2; 37576_at is PCP4
IF (36931_at < 418_at) THEN MET	36931_at is TAGLN; 418_at is MKI67
IF (774_g_at <= 1232_s_at) THEN MET	774_g_at is MYH11; 1232_s_at is IGFBP1
IF (1276_g_at < 32120_at) THEN MET	1276_g_at is RBPMS; 32120_at is SPAG5
IF (32582_at < 38430_at) THEN MET	32582_at is MYH11; 38430_at is FABP4
IF (37407_s_at < 1648_at) THEN MET	37407_s_at is MYH11; 1648_at is OSMR
IF (773_at <= 32306_g_at) THEN MET	773_at is MYH11; 32306_g_at is COL1A2
IF (39714_at <= 1260_s_at) THEN MET	39714_at is SH3BGRL; 1260_s_at is GSTA1
IF (39109_at >= 828_at) THEN MET	39109_at is TPX2; 828_at is PTGER2
IF (36572_r_at > 37407_s_at) THEN MET	36572_r_at is ARL6IP1; 37407_s_at is MYH11
IF (774_g_at < 35970_g_at) THEN MET	774_g_at is MYH11; 35970_g_at is MPHOSPH9
IF (767_at <= 546_at) THEN MET	767_at is MYH11; 546_at is PKIA
IF (773_at <= 37749_at) THEN MET	773_at is MYH11; 37749_at is MEST
IF (418_at > 36931_at) THEN MET	418_at is MKI67; 36931_at is TAGLN
IF (32272_at >= 32582_at) THEN MET	32272_at is TUBA3; 32582_at is MYH11
<b>C. Classifiers that distinguish primary prostate cancer from benign control</b>	
IF (IMAGE:510856 + IMAGE:288663) >= -0.890150 THEN PCA	IMAGE:510856 is ENC1; IMAGE:288663 is GJB1
IF (IMAGE:470216 + IMAGE:788180) >= -2.677600 THEN PCA	IMAGE:470216 is MYO6; IMAGE:788180 is AMACR
IF (IMAGE:416374 + IMAGE:244391) >= -0.417200 THEN PCA	IMAGE:416374 is TSPAN13; IMAGE:244391 is PRKCBP1
IF (IMAGE:460164 + IMAGE:2043415) >= -0.776500 THEN PCA	IMAGE:460164 is C20ORF74; IMAGE:2043415 is DAPK1
IF (IMAGE:396839 + IMAGE:510856) >= -0.551250 THEN PCA	IMAGE:396839 is UniGene Hs.601168, transcribed locus; IMAGE:510856 is ENC1
<b>D. Classifiers that distinguish high-risk lung adenocarcinomas from low-risk ones</b>	
IF (Z37976_at - D28473_s_at) <= (D14874_at + (U91327_at * X68090_s_at)) THEN High-Risk	Z37976_at is LTBP2; D28473_s_at is IARS; D14874_at is ADM; U91327_at is CCT2; X68090_s_at is FCGR2A;
IF (M29610_at - Z70218_s_at) < (D14874_at + (M23161_at + X54942_at)) THEN High-Risk	M29610_at is GYPB; Z70218_s_at is MN1; D14874_at is ADM; M23161_at is MCFD2; X54942_at is CKS2;
IF (D87465_at - U57721_at) < (HG4114-HT4384_at - (D26155_s_at - D21090_at)) THEN High-Risk	D87465_at is SPOCK2; U57721_at is KYNU; HG4114-HT4384_at is unknown transcript; D26155_s_at is SMARCA2; D21090_at is RAD23B;
IF (L38486_at - M63180_at) <= (X68090_s_at - (D31888_at * D14874_at)) THEN High-Risk	L38486_at is MFAP4; M63180_at is TARS; X68090_s_at is FCGR2A; D31888_at is RCOR1; D14874_at is ADM;
IF (L24804_at + L20320_at) >= (U70867_at + (L38486_at / Z49254_at)) THEN High-Risk	L24804_at is PTGES3; L20320_at is CDK7; U70867_at is SLCO2A1; L38486_at is MFAP4; Z49254_at is MRPL23;